Automated Evaluation Metric for Terminology Consistency in MT

Kirill Semenov, Ondrej Bojar
kir.semenow@yandex.ru; bojar@ufal.mff.cuni.cz
Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, Czech Republic

Introduction

- For some domain (such as legal texts), term consistency is crucial
- The proper NMT system should meet three requirements:
  - **"consistency": 1 source term -> 1 target term**
  - **"unambiguity": 1 target term -> 1 source term**
  - "adequacy": each target term is adequate translation of source term

We concentrate on those two parameters

Metric

STEP 1: Source term extraction
reduces to keyword extraction:
- automated solutions (YAKE, RAKE, KeyBERT, ...)
- semi-manual ad hoc solutions (regex for term introductions)

STEP 2: Term alignment
reduces to automated word alignment:
- unsupervised methods (fasttext, MOSES, ...)

STEP 3: Choosing "pseudo-reference" term translations
which target term is "correct"?
- firstly occurred
- most frequent

Src sentence	Tgt sentence	Src terms	Tgt terms	Pseudo-ref terms
... jako nájemkyně ... (dále jen " nájemkyně ") ...	... as a tenant ... (hereinafter referred to as the " tenant ") ...	nájemkyně; nájemkyně	tenant; tenant	tenant; tenant
„ Nájem " znamená , že ...	" hiring " means that ...	Nájem	hiring	lease
Tento Dodatek č . 1 ...	this appendix no 1 ...	Dodatek č . 1	appendix no 1	appendix no 1

Evaluation - usual ML metrics:
F1 score
% of correct occurrences for each term

WMT'22 System Ranking

- ELITR corpus (legal domain), 33 CS->EN texts
- Pairwise Kendall's Tau correlations between our metric (different setups) and the standard metrics:

Metrics Compared	τ 2021	τ 2022
1st;F1 VS BLEU	0.357	-0.527
1st;F1 VS chrf	0.286	
1st;F1 VS DA	0.714*	N/A
1st;Own VS BLEU	0.143	-0.636
1st;Own VS chrf	0.071	
1st;Own VS DA	0.5	N/A
Freq;F1 VS BLEU	0.143	-0.527
Freq;F1 VS chrf	0.071	
Freq;F1 VS DA	0.786*	N/A
Freq;Own VS BLEU	-0.071	-0.636
Freq;Own VS chrf	-0.143	
Freq;Own VS DA	0.571	N/A

Discussion

1. All 3 steps of preprocessing rely on (semi-)machine algorithms
   a. regex contexts are domain- and document-dependent
2. Linguistic problems:
   a. what is a term?
   b. coreference, term synonyms:

X_j, hereinafter referred to as "Seller$_j$", and Y_k, hereinafter referred to as "Buyer$_k$", together also as "contracting parties$_{j+k}$"...

github repo with code and paper

This work was supported by GA ˇCR EXPRO grantLUSyD (20-16819X, RIV: GX20-16819X) and weused services provided by the LINDAT/CLARIAH-CZ Research Infrastructure (https://lindat.cz), supported by the Ministry of Education, Youthand Sports of the Czech Republic (Project No.LM2018101).